



RAW SEQUENCE LISTING
ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/933,780B
Source: Dipe
Date Processed by STIC: 2/14/03

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:
U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7th Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202
Or
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 01/29/2002

apl

Raw Sequence Listing Error Summary

<u>ERROR DETECTED</u>	<u>SUGGESTED CORRECTION</u>	SERIAL NUMBER: <u>091933, 7808</u>
ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE		
1 _____ Wrapped Nucleics Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."	
2 _____ Invalid Line Length	The rules require that a line not exceed 72 characters in length. This includes white spaces.	
3 _____ Misaligned Amino Numbering	The numbering under each 5 th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.	
4 _____ Non-ASCII	The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.	
5 _____ Variable Length	Sequence(s) _____ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.	
6 _____ PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.	
7 _____ Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.	
8 _____ Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence. <210> sequence id number <400> sequence id number 000	
9 _____ Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.	
10 _____ Invalid <213> Response	Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence	
11 <u>✓</u> _____ Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)	
12 _____ PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.	
13 _____ Misuse of n	n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.	



Dyp

Does Not Comply
Corrected Diskette Needed

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003
TIME: 07:42:45

Errors on pp. 3-5

Input Set : A:\HMR2053USNP1sqli.txt
Output Set: N:\CRF4\02142003\I933780B.raw

```

3 <110> APPLICANT: AVENTIS PHARMACEUTICALS INC.
4      GUO, Yong
5      MORSE, Clarence
6      YAO, Zhengbin
8 <120> TITLE OF INVENTION: MEMBRANE PENETRATING PEPTIDES AND USES THEREOF
10 <130> FILE REFERENCE: HMR2053 USNP1
12 <140> CURRENT APPLICATION NUMBER: 09/933,780B
13 <141> CURRENT FILING DATE: 2001-08-21
15 <150> PRIOR APPLICATION NUMBER: US 60/27,647
16 <151> PRIOR FILING DATE: 2000-08-25
18 <150> PRIOR APPLICATION NUMBER: GB 0103110.3
19 <151> PRIOR FILING DATE: 2001-02-07
21 <160> NUMBER OF SEQ ID NOS: 54
23 <170> SOFTWARE: PatentIn version 3.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 10
27 <212> TYPE: PRT
28 <213> ORGANISM: Artificial
30 <220> FEATURE:
31 <223> OTHER INFORMATION: Sequence of nuclear location sequence contained within the N-
term
32      inal of IL-alpha propiece
34 <400> SEQUENCE: 1
36 Asn Gly Lys Val Leu Lys Lys Arg Arg Leu
37 1          5          10
39 <210> SEQ ID NO: 2
40 <211> LENGTH: 16
41 <212> TYPE: PRT
42 <213> ORGANISM: Artificial
44 <220> FEATURE:
45 <223> OTHER INFORMATION: Signal sequence peptide from Antennapedia homeodomain
47 <400> SEQUENCE: 2
49 Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys
50 1          5          10          15
52 <210> SEQ ID NO: 3
53 <211> LENGTH: 15
54 <212> TYPE: PRT
55 <213> ORGANISM: Artificial
57 <220> FEATURE:
58 <223> OTHER INFORMATION: The fibroblast growth factor signal sequence peptide
60 <400> SEQUENCE: 3
62 Ala Ala Val Ala Leu Leu Pro Ala Val Leu Leu Ala Leu Leu Ala
63 1          5          10          15
65 <210> SEQ ID NO: 4

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003

TIME: 07:42:45

Input Set : A:\HMR2053USNP1sqli.txt

Output Set: N:\CRF4\02142003\I933780B.raw

66 <211> LENGTH: 29
67 <212> TYPE: PRT
68 <213> ORGANISM: Artificial
70 <220> FEATURE:
71 <223> OTHER INFORMATION: HIV tat signal sequence peptide
73 <400> SEQUENCE: 4
75 Cys Phe Ile Thr Lys Ala Leu Gly Ile Ser Tyr Gly Arg Lys Lys Arg
76 1 5 10 15
78 Arg Gln Arg Arg Arg Pro Pro Gln Gly Ser Gln Thr His
79 20 25
81 <210> SEQ ID NO: 5
82 <211> LENGTH: 4
83 <212> TYPE: PRT
84 <213> ORGANISM: Artificial
86 <220> FEATURE:
87 <223> OTHER INFORMATION: Peptide sequence of an N-terminal fluorescein isothiocyanate
(FIT
88 C) peptide moti
90 <400> SEQUENCE: 5
92 Gly Gly Gly Gly
93 1
95 <210> SEQ ID NO: 6
96 <211> LENGTH: 7
97 <212> TYPE: PRT
98 <213> ORGANISM: Artificial
100 <220> FEATURE:
101 <223> OTHER INFORMATION: Fragment of IFN-gamma
103 <400> SEQUENCE: 6
105 Arg Lys Arg Lys Arg Ser Arg
106 1 5
108 <210> SEQ ID NO: 7
109 <211> LENGTH: 7
110 <212> TYPE: PRT
111 <213> ORGANISM: Artificial
113 <220> FEATURE:
114 <223> OTHER INFORMATION: Frgament of the N-terminus of fibroblast growth factor.
116 <400> SEQUENCE: 7
118 Asn Tyr Lys Lys Pro Lys Leu
119 1 5
121 <210> SEQ ID NO: 8
122 <211> LENGTH: 8
123 <212> TYPE: PRT
124 <213> ORGANISM: Artificial
126 <220> FEATURE:
127 <223> OTHER INFORMATION: Luinus luteus nuclear protein import sequence
129 <400> SEQUENCE: 8
131 Lys Pro Lys Lys Lys Lys Glu Lys
132 1 5
134 <210> SEQ ID NO: 9
135 <211> LENGTH: 5

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003

TIME: 07:42:45

Input Set : A:\HMR2053USNP1splt.txt

Output Set: N:\CRF4\02142003\I933780B.raw

136 <212> TYPE: PRT
 137 <213> ORGANISM: Artificial
 139 <220> FEATURE:
 140 <223> OTHER INFORMATION: Sequence of the basic motif in the nuclear protein import
 sequenc
 141 e of Smad 3 protei
 143 <400> SEQUENCE: 9
 145 Lys Lys Leu Lys Lys
 146 1 5
 148 <210> SEQ ID NO: 10
 149 <211> LENGTH: 11
 150 <212> TYPE: PRT
 151 <213> ORGANISM: Artificial
 153 <220> FEATURE:
 154 <223> OTHER INFORMATION: Sequence of intracellular loop of 5HT2A receptor
 156 <400> SEQUENCE: 10
 158 Ser Leu Glu Lys Lys Leu Gln Asn Ala Thr Asn
 159 1 5 10
 161 <210> SEQ ID NO: 11
 162 <211> LENGTH: 23
 163 <212> TYPE: PRT
 164 <213> ORGANISM: Artificial
 166 <220> FEATURE:
 167 <223> OTHER INFORMATION: Sequence of C-terminal transmembrane 7 domain derived from
 5HT2A
 168 recepto
 170 <400> SEQUENCE: 11
 172 Lys Thr Tyr Arg Ser Ala Phe Ser Arg Tyr Ile Gln Tyr Lys Glu Asn
 173 1 5 10 15
 175 Lys Lys Pro Leu Gln Leu Ile
 176 20
 178 <210> SEQ ID NO: 12
 179 <211> LENGTH: 9
 180 <212> TYPE: PRT
 181 <213> ORGANISM: Artificial
 183 <220> FEATURE:
 184 <223> OTHER INFORMATION: Fragment of HIV TAT
 186 <400> SEQUENCE: 12
 188 Arg Lys Lys Arg Arg Gln Arg Arg Arg
 189 1 5
 191 <210> SEQ ID NO: 13
 192 <211> LENGTH: 4
 193 <212> TYPE: PRT
 194 <213> ORGANISM: Artificial
 196 <220> FEATURE:
 197 <223> OTHER INFORMATION: (peptide) - must explain genetic source, e.g. "synthetic," etc
 199 <400> SEQUENCE: 13
 201 Gly Phe Leu Gly
 202 1
 204 <210> SEQ ID NO: 14
 205 <211> LENGTH: 5

The type of errors shown exist throughout the Sequence Listing. Please check subsequent sequences for similar errors.

RAW SEQUENCE LISTING

DATE: 02/14/2003

PATENT APPLICATION: US/09/933,780B

TIME: 07:42:45

Input Set : A:\HMR2053USNP1splt.txt

Output Set: N:\CRF4\02142003\I933780B.raw

206 <212> TYPE: PRT
207 <213> ORGANISM: Artificial
209 <220> FEATURE:
210 <223> OTHER INFORMATION: Peptide
212 <400> SEQUENCE: 14
214 Asp Asp Asp Asp Lys
215 1 5
217 <210> SEQ ID NO: 15
218 <211> LENGTH: 4
219 <212> TYPE: PRT
220 <213> ORGANISM: Artificial
222 <220> FEATURE:
223 <223> OTHER INFORMATION: peptide
225 <400> SEQUENCE: 15
227 Glu Tyr Phe Pro
228 1
230 <210> SEQ ID NO: 16
231 <211> LENGTH: 16
232 <212> TYPE: PRT
233 <213> ORGANISM: Artificial
235 <220> FEATURE:
236 <223> OTHER INFORMATION: Nuclear protein import sequence of hPER1
238 <400> SEQUENCE: 16
240 Ser Arg Arg His His Cys Arg Ser Lys Ala Lys Arg Ser Arg His His
241 1 5 10 15
243 <210> SEQ ID NO: 17
244 <211> LENGTH: 16
245 <212> TYPE: PRT
246 <213> ORGANISM: Artificial
248 <220> FEATURE:
249 <223> OTHER INFORMATION: Peptide
251 <400> SEQUENCE: 17
253 Gly Arg Arg His His Cys Arg Ser Lys Ala Lys Arg Ser Arg His His
254 1 5 10 15
256 <210> SEQ ID NO: 18
257 <211> LENGTH: 23
258 <212> TYPE: PRT
259 <213> ORGANISM: Artificial
261 <220> FEATURE:
262 <223> OTHER INFORMATION: peptide
264 <400> SEQUENCE: 18
266 Gly Met Asp Tyr Lys Asp Asp Asp Asp Lys Gly Tyr Gly Arg Lys Lys
267 1 5 10 15
269 Lys Arg Arg Gln Arg Arg Arg
270 20
272 <210> SEQ ID NO: 19
273 <211> LENGTH: 23
274 <212> TYPE: PRT
275 <213> ORGANISM: Artificial

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003

TIME: 07:42:45

Input Set : A:\HMR2053USNP1sqli.txt

Output Set: N:\CRF4\02142003\I933780B.raw

277 <220> FEATURE:
278 <223> OTHER INFORMATION: peptide
280 <400> SEQUENCE: 19
282 Gly Met Asp Tyr Lys Asp Asp Asp Asp Lys Gly Tyr Gly Arg Lys Lys
283 1 5 10 15
285 Lys Arg Arg Gln Arg Arg Arg
286 20
288 <210> SEQ ID NO: 20
289 <211> LENGTH: 19
290 <212> TYPE: PRT
291 <213> ORGANISM: Artificial
293 <220> FEATURE:
294 <223> OTHER INFORMATION: peptide
296 <400> SEQUENCE: 20
298 Gly Met Asp Tyr Lys Asp Asp Asp Asp Lys Gly Met Asp Tyr Asp Asp
299 1 5 10 15
301 Asp Asp Lys
304 <210> SEQ ID NO: 21
305 <211> LENGTH: 17
306 <212> TYPE: PRT
307 <213> ORGANISM: Artificial
309 <220> FEATURE:
310 <223> OTHER INFORMATION: peptide
312 <400> SEQUENCE: 21
314 Gly Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys
315 1 5 10 15
317 Lys
320 <210> SEQ ID NO: 22
321 <211> LENGTH: 10
322 <212> TYPE: PRT
323 <213> ORGANISM: Artificial
325 <220> FEATURE:
326 <223> OTHER INFORMATION: peptide
328 <400> SEQUENCE: 22
330 Gly Arg Arg Arg Arg Arg Arg Arg Arg
331 1 5 10
333 <210> SEQ ID NO: 23
334 <211> LENGTH: 10
335 <212> TYPE: PRT
336 <213> ORGANISM: Artificial
338 <220> FEATURE:
339 <223> OTHER INFORMATION: peptide
341 <400> SEQUENCE: 23
343 Gly Lys Lys Lys Lys Lys Lys Lys Lys
344 1 5 10
346 <210> SEQ ID NO: 24
347 <211> LENGTH: 10
348 <212> TYPE: PRT
349 <213> ORGANISM: Artificial

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003
TIME: 07:42:46

Input Set : A:\HMR2053USNP1splt.txt
Output Set: N:\CRF4\02142003\I933780B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:54; Xaa Pos. 1,2,3,4

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27
Seq#:28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51
Seq#:52,53,54

VARIABLE LOCATION SUMMARY

DATE: 02/14/2003

PATENT APPLICATION: US/09/933,780B

TIME: 07:42:46

Input Set : A:\HMR2053USNP1splt.txt

Output Set: N:\CRF4\02142003\I933780B.raw

Use of n's or Xaa's (NEW RULES):

Use of n's and/or Xaa's have been detected in the Sequence Listing.

Use of <220> to <223> is MANDATORY if n's or Xaa's are present.

in <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.

Seq#:54; Xaa Pos. 1,2,3,4

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/933,780B

DATE: 02/14/2003

TIME: 07:42:46

Input Set : A:\HMR2053USNP1sqli.txt

Output Set: N:\CRF4\02142003\I933780B.raw

L:772 M:258 W: Mandatory Feature missing, <222> Tag not found for SEQ ID#:54
L:772 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:54 after pos.:0